

## **REMARKS**

In the Office Action, the Examiner noted that Claims 1 through 13 were pending in the Application. The Examiner rejected Claims 1 - 13. Applicants traverse the rejections below.

### **I. Traversal of the Rejections over the Cited Art**

The Examiner rejected Claims 1 - 13 under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent Number 6,507,840 to Ioannidis et al (Ioannidis) in view of U.S. Patent Number 6,636,862 to Lundahl et al (Lundahl). Applicants traverse these rejections below.

#### **A. The Present Invention**

The present invention provides a technique, in a computer environment, for determining an objective quality index for the result of a clustering operation. Data clustering aims to group records in a large set into clusters such that records belonging to the same cluster have a high degree of similarity. Each cluster has a set of buckets for each variable. According to one aspect of the invention, a foreground frequency is then determined for a bucket, and a background frequency is determined for the bucket with respect to all of the clusters. The foreground and background frequencies are compared, and a quality index is determined based on the comparison.

#### **B. Differences between the Claims and the Cited Art**

Ioannidis discloses a method for generating an approximate answer in response to a query to a database. An SQL quest Q for operating on a relation R in a database is received. Relation

R has an associated histogram H. The SQL query Q is translated to be a query Q' for operating on histogram H. Translated query Q' is executed on histogram H for obtaining a result histogram. The result histogram is expanded into a relation having tuples containing approximate attribute values. (See Abstract). Further, Ioannidis states that “[i]n principle, the concepts of using histograms for obtaining approximate answers can be generalized for much more complex queries involving multiple attributes for the same relation...” (Column 6, Lines 3-6). “After all, histograms maintained within a database are an approximation of the database, so executing a query on the histograms provides an approximation of the actual answer.” (Column 6, Lines 14-17). “Identifying an appropriate metric for measuring the distance between two multisets is essential for any systematic study of approximation of set-valued query answers. In the case of the present invention, the distance that is to be measured is between the multiset representing the actual answer and the multiset representing the approximate answer. (Column 6, Lines 33-40). Ioannidis then goes on to discuss approaches for comparing sets, and the Office Action applies this discussion to the third element of Claim 1 of the present invention.

Clearly, Ioannidis uses histograms in order to improve database queries. In particular, they are used to compute distances between two sets. Ioannidis is not related to data clustering or other data mining techniques. Ioannidis apparently uses the term ‘clustering’ only once, and that is in conjunction with the Hausdorff distance, which “addresses aspects of set differences that are different from approximate query answering.” (Column 6, lines 40-46).

So, clearly, Ioannidis is not directed to clustering, or determining the quality of the result of a clustering data processing operation. Further, Ioannidis does not address, teach, suggest or disclose a foreground frequency or a background frequency.

Claim 1 recites “determining a foreground frequency of a bucket within a first cluster”. Relative to this subject matter, the Office Action cites Column 8, Lines 62-67 and Column 9, Lines 1-30. This passage discusses definitions “which are useful for understanding the histogram-based approximate query answering technique of [Ioannidis].” (Column 8, Lines 64-

66). The passage, in defining a histogram, talks about a partitioning rule for partitioning the data distribution of an attribute X “into mutually disjoint subsets called buckets and approximating the frequencies and values in each bucket in some well-known manner.” The terms ‘frequency’ and ‘bucket’ are employed in this passage. But there is no use of these terms in any manner that teaches, suggests or discloses the present invention. No foreground frequency is determined. No foreground frequency of a bucket within a first cluster is determined. Note that Ioannidis itself has nothing to do with clusters, but rather approximate answers to data queries.

Claim 1 further recites “determining a background frequency of the bucket with respect to all of the clusters”. Relative to this subject matter, a passage from Column 10, Lines 1-30 is cited. This passage addresses the accuracy of the approximation. Different types of histograms are discussed. The use of buckets in histograms is discussed. But there is no teaching, suggestion or disclosure of “determining a background frequency”. There is no discussion of a background frequency in Ioannidis. There is no teaching, suggestion or disclosure of “determining a background frequency of the bucket with respect to all of the clusters”. No clusters are discussed or described at all in Ioannidis.

Claim 1 further recites “comparing the foreground and background frequencies”. Relative to this subject matter, the Office Action cites a passage from Column 6, lines 52-67. This passage, as discussed above, describes “[t]wo well-known approaches for comparing sets”. There is no discussion of a foreground frequency or a background frequency. There is a description of equation elements that represent the “frequency sets of two data distributions on a universe of  $n$  elements”. But this does not teach, suggest or disclose the cited subject matter from Claim 1.

In summary, some terms that are employed in Claim 1 are also found in Ioannidis. It is not even clear that the terms are used in a way such that they even represent the same things. But the functionality recited in Claim 1 is in no way taught, suggested or disclosed by Ioannidis. The problem solved by the Ioannidis invention is completely different from the problem solved by

the present invention. The functionality is completely different. The assertion that Ioannidis has teaches anything with respect to data clustering has no support in Ioannidis. Ioannidis is a histogram-based technique for providing approximate answers in response to a database query. The use of similar terms does not render the subject matter of Claim 1 obvious.

Lundahl describes a technique for the dynamic analysis of data. Lundahl has no similarity to a histogram-based technique for providing approximate answers in response to a database query as described in Ioannidis. Lindahl does mention clusters. But how Lindahl are combined to teach the subject matter of Claim 1 is unclear. Lindahl was not applied against the elements of Claim 1 discussed above. These elements are clearly not taught, suggested or disclosed by Ioannidis, and no suggestion is made in the Office Action that they are taught, suggested or disclosed by Lindahl.

Accordingly, Applicants submit that independent Claim 1 patentably distinguishes over the combination of Ioannidis and Lindahl. It follows that dependent Claims 2 - 9 also distinguish therefrom. And since independent Claim 13 was rejected for the same reasons as was Claim 1, it follows that Claim 13 also distinguishes over the combination.

Independent Claim 10 recites “performing a number of data clustering operations”. Relative to this subject matter, the Office Action cites passages from Columns 12 and 4 of Ioannidis. Neither of these passages has anything to do with data clustering. As discussed above, Ioannidis is directed to generating an approximate answer in response to a query to a database. The present invention has to do with determining an objective quality index for the result of a clustering operation. The Office Action appears to be suggesting that these two are the same functionality with respect to the arguments applied to Claim 10.

Independent Claim 10 also recites “determining a quality index for each result of the data clustering operations”. First, Ioannidis performs no data clustering. One type of metric discussed in Ioannidis, the Hausdorff distance, “is used in connection with data clustering.”

(Column 6, Lines 40-45). This does not mean that Ioannidis performs data clustering; rather, it calculates approximate answers in response to queries. Accordingly, Ioannidis does not teach, suggest or disclose the second step of Claim 10.

Accordingly, Applicants submit that since Ioannidis does not in fact teach, suggest or disclose those portions of Claim 10 it is applied against, Claim 10 patentably distinguishes over the combination of Ioannidis and Lundahl.

Independent Claim 11 recites “a method for data clustering....comprising....selecting an initial set of clusters. Relative to this subject matter, a passage from Column 13 is cited, and the Ioannidis selection of an initial element in the bucket is cited. First, as discussed above, Ioannidis does not perform data clustering. Second, selection of an initial element from a bucket does not teach, suggest or disclose selecting an initial set of clusters. Further, relative to Claim 1, it was apparently alleged that the bucket from Ioannidis taught a bucket from Claim 1. This position has apparently changed, so that the bucket from Ioannidis now has to do with clusters, and that an element in the bucket teaches a set of clusters. These positions are contradictory.

Accordingly, Applicants submit that since Ioannidis does not in fact teach, suggest or disclose those portions of Claim 11 to which it is applied, Claim 11 patentably distinguishes over the combination of Ioannidis and Lundahl. It follows that dependent Claim 12 also distinguishes therefrom.

### C. Improper Combination of References

Applicants also submit that the combination of references is improper. There must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. The

teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, and not based on applicant's disclosure. *In re Vaeck*, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991). See MPEP 2143 - 2143.03 for decisions pertinent to each of these criteria.

The initial burden is on the examiner to provide some suggestion of the desirability of doing what the inventor has done. "To support the conclusion that the claimed invention is directed to obvious subject matter, either the references must expressly or impliedly suggest the claimed invention or the examiner must present a convincing line of reasoning as to why the artisan would have found the claimed invention to have been obvious in light of the teachings of the references." *Ex parte Clapp*, 227 USPQ 972, 973 (Bd. Pat. App. & Inter. 1985). See MPEP Section 2144 - Section 2144.09 for examples of reasoning supporting obviousness rejections.

When the motivation to combine the teachings of the references is not immediately apparent, it is the duty of the examiner to explain why the combination of the teachings is proper. *Ex parte Skinner*, 2 USPQ2d 1788 (Bd. Pat. App. & Inter. 1986).


Applicants submit that the criteria described above for combining the references has not been met. No convincing line of reasoning was provided for combining the technique for dynamic analysis of data from Lundahl with the technique for histogram-based approximation for generating approximate answers in response to a query to a database from Ioannidis. Nothing from either reference suggested their combination. Accordingly, Applicants submit that the combination is improper.

### III. Summary

Applicants have presented technical explanations and arguments fully supporting their

position that the pending claims contain subject matter which is not taught, suggested or disclosed by Ioannidis, Lundahl, or any combination thereof. Accordingly, Applicants submit that the present Application is in a condition for Allowance. Reconsideration of the claims and a Notice of Allowance are earnestly solicited.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Greg M. Doudnikoff", written in a cursive style.

Gregory M. Doudnikoff  
Attorney for Applicant  
Reg. No. 32,847

GMD/ssc

Docket No: DE920000057US1

PHONE: 919-254-1288 FAX: 919-254-4330